A mutation in the 5′ untranslated region of the human α-galactosidase A gene in high-activity variants inhibits specific protein binding

Zubaida Saifudeen[a], Robert J. Desnick[b], Melanie Ehrlich[a,c,*]

[a]*Molecular and Cellular Biology Program, Mount Sinai School of Medicine, New York, NY 10029, USA*
[b]*Department of Medical and Molecular Genetics, Mount Sinai School of Medicine, New York, NY 10029, USA*
[c]*Department of Biochemistry, Tulane Medical School, New Orleans, LA 70112, USA*

Received 22 July 1995

Abstract Recently, normal individuals were identified who had high levels of plasma α-galactosidase A activity and a G to A transition in the 5′ untranslated (5′ UT) region of the α-galactosidase A gene. Electrophoretic mobility shift assays revealed that the wild-type sequence at the site of this mutation complexed with specific nuclear proteins. A standard NF-κB site competed with the 5′ UT site for formation of these DNA–protein complexes. Complex formation was inhibited by the transition mutation. Therefore, the wild-type site might down-modulate expression of the α-galactosidase A gene from this 5′ untranslated region, which includes a previously described protein-binding site for another family of sequence-specific DNA-binding proteins, methylated DNA-binding protein.

Key words: Transcription factor; α-Galactosidase A; Promoter sequence; NF-κB; DNA methylation

1. Introduction

Human α-galactosidase A (EC 3.2.1.22; α-Gal A), a lysosomal exoglycosidase, is encoded by a single X-linked 12-kb housekeeping gene, which has been cloned and sequenced along with the corresponding 5′ flanking region and 1.4-kb cDNA [1–4]. This enzyme is present in a variety of human tissues [5] and a study of rats showed that its levels vary greatly from organ to organ [6]. A severe deficiency in α-Gal A activity results in Fabry disease, a glycosphingolipid storage disorder. Many different mutations in the *α-GAL A* gene have been correlated with this disease, including partial gene deletions, insertions, and a variety of single-base substitutions affecting mRNA processing or translation [7,8].

The major transcription start point of the *α-GAL A* gene is about 60 bp upstream of the starting codon [9]. Although the promoter region has not yet been dissected, the immediate upstream region of this gene contains several motifs similar to those recognized by sequence-specific DNA-binding proteins involved in transcription control as well as a cluster of repeated 10-bp sequences of unknown specificity [9]. Like many other housekeeping genes, this region is rich in CpG dinucleotides compared to the overall CpG content of mammalian DNA.

During a recent screening of about 400 individuals, a G to A transition in the 5′ untranslated (5′ UT) region of the *α-GAL A* gene was identified in several normal individuals who had plasma α-Gal A activity levels that were 2- to 7-fold higher than the mean activity in random plasma samples [10]. The high-activity phenotype was inherited as an X-linked trait affecting only this lysosomal enzyme. No mutation other than this G to A transition was seen upon sequencing the full-length *α-GAL A* cDNA, all eight intron–exon boundaries of the gene, and almost 200 bp of the upstream region from a high-activity variant. We report here that this mutation inhibited specific protein binding to the 5′ UT sequence of the *α-GAL A* gene and might, thereby, be responsible for the high activity level associated with this variant allele.

2. Materials and methods

2.1. Oligonucleotides

Oligonucleotides were synthesized, labeled with [γ-^{32}P]ATP, and annealed to form blunt-ended duplexes as described previously [11]. In addition to the oligonucleotides shown in Fig. 1, the following duplexes (written 5′ to 3′) were used, with only one strand of the sequence given and the MDBP or AP-2 sites in capital letters: EIA1, tttaGTTGCCTAGCAACatga; Py2, ggcaGTTGCCTAGCAACtaat; Etp, tcatGTCTCCATGGTAACttac (containing MDBP binding sites from the equine infectious anemia virus long terminal repeat, the polyomavirus enhancer B, and the *ets*-2 promoter, respectively; see [12,13] and unpublished results) and Ap-2, gggagcaaagGCCTGGGGTgca (see [14] and unpublished results).

2.2. Nuclear extracts

Nuclear extracts were prepared as previously described [12] from HeLa or Raji cells grown in medium containing 10% fetal bovine serum except that the HeLa cells, where indicated, had been treated with 5 μM forskolin and 100 ng/ml of phorbol 12-myristate 13-acetate (PMA) for 1 h and then harvested.

2.3. EMSA

Analysis of DNA–protein complex formation by electrophoretic mobility shift assay(s) (EMSA) was conducted as previously described with 0.3–2 μl of nuclear extracts (~3–6 mg of protein per ml) pre-incubated with 1 μg of poly(dI-dC)·poly(dI-dC), as a non-specific competitor, for 10 min followed by a 15-min incubation at room temperature with 20–40 fmol of ^{32}P-labeled ligand and then electrophoresis on a 5% polyacrylamide gel [15]. When a specific unlabeled oligonucleotide duplex competitor was present (2 pmol unless otherwise noted), it was added 10 min before the radiolabeled ligand. Radiolabeled complexes were visualized upon autoradiography and quantitated by phosphorimager analysis.

*Corresponding author. Department of Biochemistry/SL43, Tulane Medical School, 1430 Tulane Ave., New Orleans, LA 70112, USA. Fax: (1) (504) 584-2739.

Abbreviations: α-Gal A, α-galactosidase A; WT, wild-type; EMSA, electrophoretic mobility shift assay; kb, kilobase-pair; bp, base-pair; 5′ UT, 5′ untranslated; Lwt, specific protein binding site in the 5′ UT region of the WT *α-GAL A* gene; Lmu, mutant form of the Lwt site.

SSDI 0014-5793(95)00891-8

```
5' UT α-GAL A        5'-GGTTACCCGCGGAAATTTATGCTGTCCGGTCACCGTGACAATG-3'
Lwt                  5'-GGTTACCCGCGGAAATTTATG-3'
Lmu                  5'-GGTTACCCGCaGAAATTTATG-3'
NF-κB consensus            5'-GGGRNNYYCC-3'
κB                   5'-GCCAATAGGGACTTTCCATT-3'
```

Fig. 1. Part of the 5′ UT sequence of the α-*GAL A* gene and related oligonucleotide duplexes. A region of the α-*GAL A* gene immediately upstream of the initiation codon (doubly underlined) is shown (5′ UT α-*GAL A*) with the WT G residue that is mutated to an A in high-activity variants in italics and the previously demonstrated MDBP binding site [12] in boldface. The MDBP site is shown in its unmethylated form although much more binding is seen to the CpG-methylated version of this site [12]. One strand of each of the following oligonucleotide duplex ligands used for EMSA is also given: Lwt, the WT sequence from this leading portion of the α-*GAL A* gene; Lmu, the mutated form of this sequence found in several high-activity variants; and κB, a standard NF-κB ligand from the enhancer of the major immediate early gene of human cytomegalovirus [29]. The regions that match or are partially homologous to the NF-κB consensus sequence are underlined. The mutant base in Lmu is in lower case.

3. Results

3.1. *A specific protein-binding site, Lwt, in the 5′ untranslated region of the* α-GAL A *gene*

In order to test whether the α-*GAL A* gene's G to A transition, which was found in the gene's 5′ UT region in four unrelated high-activity variants [10], might affect binding of nuclear proteins, we assayed both the WT and mutant forms of this DNA sequence for specific complex formation with proteins in nuclear extracts. We noted that the WT form of the mutant sequence partially matches the NF-κB consensus sequence (Fig. 1). Therefore, we compared as ligands in EMSA the following three oligonucleotide duplexes: a consensus sequence NF-κB site, κB duplex, and the WT and mutant forms of the leader region's (5′ UT) NF-κB-like oligonucleotide sequence in the α-*GAL A* gene, Lwt or Lmu duplexes, respectively (Fig. 1).

As seen in Fig. 2, proteins in nuclear extracts from HeLa and Raji cells formed complexes with the Lwt oligonucleotide duplex banding in the region of the complexes formed with the standard κB site. The mobility of these Lwt–protein complexes was greater than that of complexes containing a family of proteins (MDBP) that binds to a previously described protein-recognition site [12] in the 5′ UT region of the α-*GAL A* gene (Fig. 3, lane 1 vs. 3 and see below). Several unlabeled oligonucleotide duplex competitors were used in excess to try to compete with the radiolabeled Lwt duplex for formation of complexes with its cognate proteins in the HeLa nuclear extract. Formation of the specific radiolabeled Lwt–protein complexes was competed by the unlabeled κB duplex as well as by an excess of unlabeled Lwt duplex, but not appreciably by binding sites for MDBP or AP-2 (Fig. 3, lanes 3, 4, 6–9). Excess κB duplex also inhibited complex formation between radiolabeled Lwt duplex and proteins in a Raji nuclear extract while this complex formation decreased minimally with the addition of an irrelevant oligonucleotide duplex (Fig. 3B, lane 5). These findings indicate that specific proteins in mammalian nuclear extracts which recognize a standard NF-κB ligand can complex with the Lwt duplex.

3.2. *Decreased protein binding to the Lmu site of the high-activity variant allele*

An oligonucleotide duplex identical to the Lwt duplex except for the G to A transition at the Lwt site was used for EMSA with nuclear extracts. Much less specific complex formed with this radiolabeled duplex (Lmu, Fig. 1) than with the corresponding WT duplex (Lwt; Fig. 2A and B). The relative extents of complex formation from Lmu vs. Lwt duplex with nuclear extracts of HeLa cells, induced HeLa cells (see below), and Raji cells were 0.2, 0.3, and 0.1, respectively. Furthermore, these ratios may be overestimates of the amount of specific complex formation by the Lmu duplex. Although the DNA–protein complexes containing the Lmu duplex were seen in the general region of those containing the Lwt duplex, the patterns of bands were different (Fig. 2A, lanes 3 vs. 2, 6 vs. 5, 9 vs. 8). Competition experiments provided further evidence of the

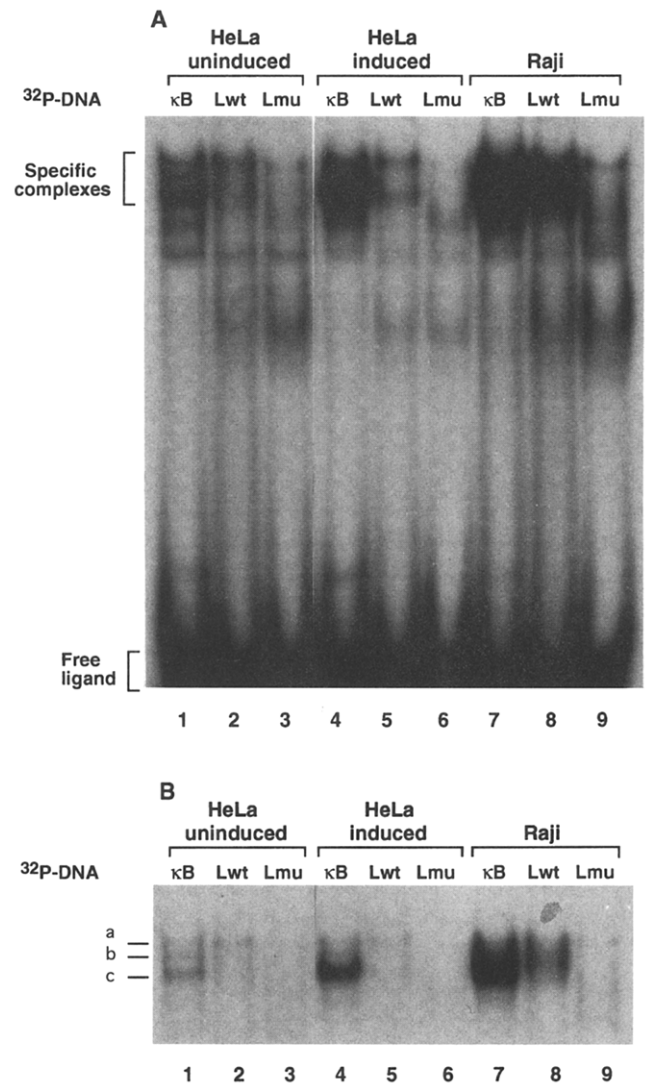

Fig. 2. Recognition of the Lwt site from the α-*GAL A* gene's 5′ UT region by DNA-binding proteins in HeLa and Raji nuclear extracts as detected by EMSA. (A) 16-h exposure for autoradiography. (B) 3-h exposure of the same gel. EMSA was done with the indicated oligonucleotide duplex (Fig. 1) and an uninduced HeLa nuclear extract, a PMA- and forskolin-induced HeLa nuclear extract, or an uninduced Raji nuclear extract using 11, 7, and 8 μg, respectively, of protein. The specific activities of the κB, Lwt, and Lmu ligands were about 4,200, 1500, and 2900 cpm/fmol, respectively, and 20 fmol of each ligand was used per assay. The three bands of specific complexes obtained with the κB duplex incubated with HeLa nuclear extracts are indicated in Panel B.

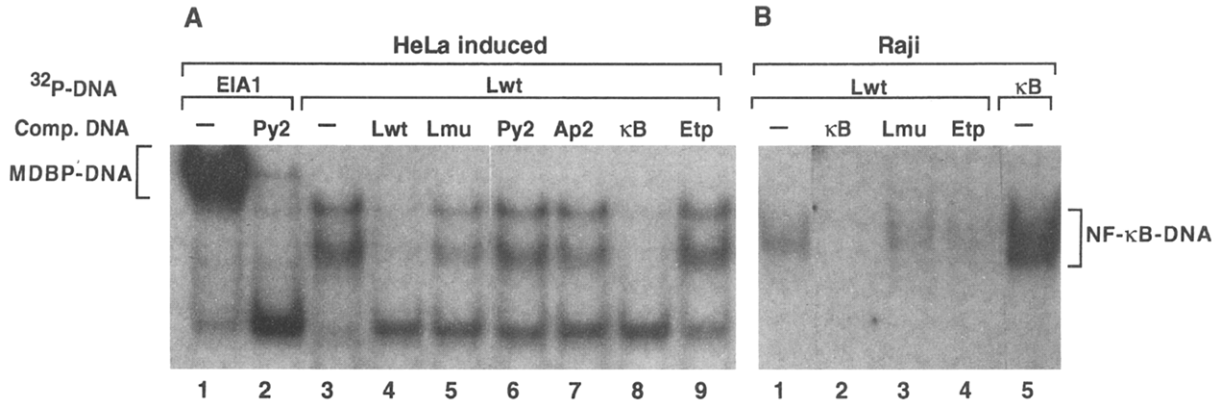

Fig. 3. Competition between various oligonucleotide duplexes for the binding of the Lwt site to proteins in nuclear extracts. (A) PMA- and forskolin-induced HeLa nuclear extract (7 μg of protein). (B) Raji nuclear extract (1 μg of protein). The sequences of the oligonucleotide duplexes used are given in Fig. 1 or section 2; Lwt is the WT NF-κB-like site in the 5′ UT region of the α-*GAL A* gene and EIA1 and Py2 are standard ligands for MDBP, which binds to a site neighboring the Lwt site in this 5′ UT region (Fig. 1). Excess, unlabeled oligonucleotide duplex was added to the reaction mixture as competitor (see section 2). Below the displayed region of the autoradiogram was one other DNA–protein complex band but its formation was not competed by the addition of any of the tested oligonucleotide duplexes, including an excess of the unlabeled form of the labeled ligand. The autoradiography exposure time for Panel B was about 1/5 the exposure time for Panel A.

lower affinity of the Lmu duplex vs. the analogous WT duplex for specific complex formation. When excess unlabeled Lmu duplex was used as a competitor in the binding reaction of radiolabeled Lwt duplex and HeLa nuclear extract proteins, less inhibition of formation of radiolabeled complex was seen than when the same excess of unlabeled Lwt duplex was added (Fig. 3A, lanes 3–5).

3.3. *Comparison of the nuclear proteins binding to the Lwt site and the major NF-κB family proteins in HeLa or Raji nuclear extracts*

Complex formation between the Lwt duplex and nuclear proteins was compared for nuclear extracts from HeLa cells grown in normal medium (uninduced extracts) and in medium supplemented with PMA and forskolin (induced extracts). Upon EMSA, the ratio of total radioactivity in specific bands from induced vs. uninduced extracts was 3.4 for the κB duplex; however, the analogous ratio for the Lwt duplex was 1.0 (Fig. 2A). Much more complex was formed with the Raji nuclear extract than with the induced or uninduced HeLa nuclear extracts when either the κB duplex or the Lwt duplex was the ligand (Fig. 2B).

Although the complexes formed between either the Lwt or κB duplexes and cognate proteins in HeLa or Raji nuclear extracts had similar electrophoretic mobilities, the patterns of specific bands upon electrophoresis were different. In nuclear extracts from uninduced or PMA- and forskolin-induced HeLa cells, a higher-mobility band (c, Fig. 2B, lanes 1 and 4) predominated over two lower-mobility bands (a and b) when the κB duplex was the ligand. With the Lwt ligand, only the a and c bands were seen and the ratio of intensities of these bands was 1.0 (Fig. 2A, lanes 2 vs. 1 and 5 vs. 4). Also, with Raji nuclear extracts, ligand-specific differences were seen; the leading edge of the band of DNA–protein complexes was lower in the gel for the κB ligand than for the Lwt duplex (Fig. 2B, lanes 7 vs. 8 and Fig. 3B, lanes 1 vs. 5).

As noted above, excess unlabeled κB duplex was a good competitor for the formation of all the specific DNA–protein complexes containing radiolabeled Lwt duplex; however, the converse was not true. With the nuclear extract from induced HeLa cells and the ^{32}P-labeled κB duplex as the ligand, 2 pmol of unlabeled Lwt duplex as the competitor decreased specific, labeled complex formation by only 38% (Fig. 4A, lanes 3 vs. 1). In the same experiment, the formation of labeled NF-κB-like complexes with the DNA ligands was decreased by 96% and 6% by the addition of 2 pmol of unlabeled κB and Lmu duplexes, respectively (Fig. 4A, lanes 2 and 4 vs. 1). Similarly, with a nuclear extract from Raji cells and radiolabeled κB duplex, 2 pmol of the unlabeled Lwt duplex competed for specific DNA–protein complex formation less efficiently than 2 pmol of unlabeled κB duplex but more effectively than the same amount of unlabeled Lmu duplex. The formation of the slower moving complexes was inhibited more than that of the faster moving complexes (Fig. 4B). Increasing the amount of unlabeled Lwt competitor to 5 pmol did not further decrease the formation of specific DNA–protein complexes between labeled κB duplex and NF-κB family proteins in the Raji nuclear extract (data not shown).

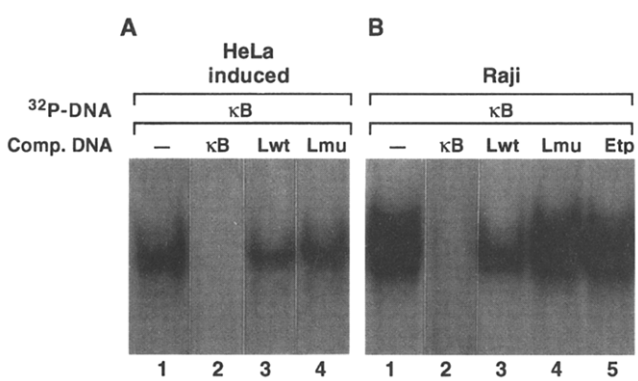

Fig. 4. The Lwt site from the α-*GAL A* 5′ UT region only partially competes with a standard NF-κB site for complex formation with nuclear proteins. (A) PMA- and forskolin-induced HeLa nuclear extract. (B) Raji nuclear extract. Competition assays were conducted as for Fig. 3 except that the exposure times for Panels A and B were approximately the same.

4. Discussion

The studies described here demonstrate that specific DNA–protein complexes are formed between proteins in nuclear extracts from Raji or HeLa cells and a sequence, designated Lwt, containing an NF-κB-like site from the 5′ UT region of the WT α-*GAL A* gene. The Lwt site deviates from the NF-κB consensus sequence by 3 bp and has a central 6-bp palindrome (Fig. 1). DNA–protein complexes containing this oligonucleotide duplex electrophoresed in the same region of the gel as those containing a standard NF-κB site (κB duplex, Fig. 1); however, ligand-dependent differences in the pattern of the DNA–protein bands were seen (Fig. 2). Also, when the Lwt duplex was used in excess to compete for formation of complexes between the κB duplex and NF-κB proteins in HeLa or Raji nuclear extracts, specific competition was seen but it was only partial (Fig. 4). In the converse experiment, the κB duplex competed efficiently for the formation of Lwt–protein complexes (Fig. 3). The Lwt-binding activity, unlike the κB duplex-binding activity in HeLa cells [16], was not inducible after a brief treatment with PMA in the presence of forskolin (Fig. 2). Therefore, the major form of NF-κB in induced HeLa nuclear extracts, p50/p65 [17,18], is not the predominant protein that bound to the Lwt sequence in the α-*GAL A* 5′ UT region. Not all NF-κB-type proteins accumulate appreciably in the nucleus upon induction, even in HeLa cells [19], and different NF-κB proteins can differ somewhat in their preferences for DNA sequences related to the NF-κB consensus sequence [20,21]. Therefore, other members of the NF-κB family, such as c-Rel-containing forms of NF-κB, which are the predominant type of NF-κB in B cell lines like Raji [17,18], might be binding to the Lwt site. Because two distinct Lwt-specific bands were seen in EMSA with HeLa nuclear extracts (Fig. 2A) and a cluster of bands were seen with Raji nuclear extracts (Fig. 2B), there may be more than one specific Lwt-binding protein.

The Lwt site in the α-*GAL A* gene's 5′ UT region identified in the present study is the location of a G to A transition recently found in six individuals displaying high plasma levels of α-Gal A activity [10]. This mutant sequence (Lmu; Fig. 1) complexed to a much lesser extent with the Lwt-specific binding activity than did the Lwt site (Figs. 2 and 3). For one of these individuals, the entire coding region and about 200 bp of the upstream region of the α-*GAL A* gene were sequenced [10]. Only this G to A transition in the 5′ UT region was seen. This mutation did not inhibit in vitro translation of the corresponding α-*GAL A* mRNA in a rabbit reticulocyte translation system using recombinant plasmid-encoded mRNA [10].

We propose that, in addition to the cap-upstream promoter, this 5′ UT gene region is important in transcription control via the binding of sequence-specific transcription factors as seen for some other Class II genes (e.g. [22,23]). The mutation associated with the high-activity α-Gal A phenotype is 19 bp upstream of a site that was shown to bind to a family of sequence-specific DNA-binding proteins called methylated DNA binding protein (MDBP; 12). For many, but not all, MDBP sites, including the α-*GAL A* site, binding is much better when the site is methylated at its CpG dinucleotides. Much evidence indicates that DNA methylation is involved in negatively controlling expression of many genes in mammals, including X-linked genes subject to lyonization [24,25]. MDBP sites have been implicated in the negative control of gene expression from certain sites as well as in positive control from others [26–28]. The 5′ UT region of the α-*GAL A* gene might be involved in down-modulation of gene expression from both the Lwt and MDBP sites, functioning either independently or co-operatively. Mutation of the Lwt site to a weaker binding site for an Lwt-specific DNA-binding protein could thereby increase the level of expression of the α-*GAL A* gene.

Acknowledgements: Supported in part by NIH Grants GM33999 to M.E. and DK34045 to R.J.D.

References

[1] Bishop, D., Calhoun, D.H., Bernstein, H.S., Hantzopoulos, P., Quinn, M. and Desnick, R.J. (1986) Proc. Natl. Acad. Sci. USA 83, 4859–4863.
[2] Tsuji, S., Martin, B.M. and Kaslow, D.C. et al. (1987) Eur. J. Biochem 165, 275–280.
[3] Bishop, D.F., Kornreich, R. and Desnick, R.J. (1988) Proc. Natl. Acad. Sci. USA 85, 3903–3907.
[4] Kornreich, R., Desnick, R.J. and Bishop, D.F. (1989) Nucleic Acids Res. 17, 3301.
[5] Bishop, D.F. and Desnick, R.J. (1981) J. Biol. Chem. 256, 1307–1316.
[6] Brady, R.O., Gal, A.E., Bradley, R.M. and Martensson, E. (1967) J. Biol. Chem. 242, 1021–1026.
[7] Bernstein, H.S., Bishop, D.F., Astrin, K.H., Kornreich, R., Eng, C.M. and Desnick, R.J. (1989) J. Clin. Invest. 83, 1390–1399.
[8] Eng, C.M. and Desnick, R.J. (1994) Hum. Mut. 3, 103–111.
[9] Quinn, M., Hantzopoulos, P., Fidanza, V. and Calhoun, D.H. (1987) Gene 58, 177–188.
[10] Fitzmaurice, T.F., Desnick, R.J. and Bishop, D.F. (1995) submitted.
[11] Khan, R., Zhang, X.-Y., Supakar, P.C., Ehrlich, K.C. and Ehrlich, M. (1988) J. Biol. Chem. 263, 14374–14383.
[12] Zhang, X.-Y., Asiedu, C.K., Supakar, P.C., Khan, R., Ehrlich, K.C. and Ehrlich, M. (1990) Nucleic Acids Res. 18, 6253–6260.
[13] Mavrothalassitis, G.J., Watson, D.K. and Papas, T.S. (1990) Proc. Natl. Acad. Sci. USA 87, 1047–1051.
[14] Ehrlich K.C. and Ehrlich, M. (1990) Mol. Cell. Biol. 10, 4957–4960.
[15] Supakar, P.C., Wiest, D., Zhang, D., Inamdar, N., Zhang, X.-Y., Ehrlich, K.C. and Ehrlich, M. (1988) Nucleic Acids Res. 16, 8029–8044.
[16] Sen, R. and Baltimore, D. (1986) Cell 47, 921–928.
[17] Liou, H.-C. and Baltimore, D. (1993) Curr. Opin. Cell Biol. 5, 477–487.
[18] Siebenlist, U., Franzoso, G. and Brown, K. (1994) Annu. Rev. Cell Biol. 10, 405–455.
[19] Naumann, M. and Scheidereit, C. (1994) EMBO J. 13, 4597–4607.
[20] Hansen, S.K., Nerlov, C., Zabel, U., Verde, P., Johnsen, M., Baeuerle, P.A. and Blasi, F. (1992) EMBO J. 11, 205–213.
[21] Kunsch, C., Rueben, S.M. and Rosen, C.A. (1992) Mol. Cell. Biol. 12, 4412–4421.
[22] Smith, K.C. and Sharp, Z.D. (1991) Biochem. Biophys. Res. Commun. 177, 790–796.
[23] Emanuel, P. and Gilmour, D.S. (1993) Proc. Natl. Acad. Sci. USA 90, 8449–8453.
[24] Martienssen, R.A. and Richards, E.J. (1995) Curr. Opin. Cell Biol. 5, 234–242.
[25] Hornstra, I.K. and Yang, T.P. (1994) Mol. Cell. Biol. 14, 1419–1430.
[26] Zhang, X.-Y., Jabrane-Ferrat, N., Asiedu, C.K., Samac, S., Peterlin, B.M. and Ehrlich, M. (1993) Mol. Cell. Biol. 13, 6810–6818.
[27] Asiedu, C.K., Scotto, L., Assoian, R.K. and Ehrlich, M. (1994) Bioch. Biophys. Acta 1219, 55–63.
[28] Siegrist, C.A., Durand, B., Emery, P., David, E., Hearing, P., Mach, B. and Reith, W. (1993) Mol. Cell. Biol. 13, 6375–6384.
[29] Sambucetti, L.C.C., Cherrington, Wilkinson, G.W.G. and Moccarski, E.S. (1989) EMBO J. 8, 4251–4258.